ÇANKAYA UNIVERSITY
FACULTY OF ENGINEERING
COMPUTER ENGINEERING DEPARTMENT

Project Report

Version 1

CENG 407

Innovative System Design and Development I

P201919

Image Analysis for the Classification of Brain Tumor Location on MR Images

Nisan Çandır
201611010
Ece Yılmaz
201611064
Kevser Sürgüveç
201411055
Deniz Yıldız
201511066

Advisor: *Faris Serdar TAŞEL*

Co-Advisor: Roya Choupani

Table of Contents

Table of Contents iii

Abstract vi

1. Introduction 1

  1.1 Problem Statement 1

  1.2 Background or Related Work 1

  1.3 Solution Statement 1

  1.4 Contribution 2

2. Literature Search 3

  2.1 Examples of Segmentation Algorithms 3

  2.2 Related Literature 8

  2.3 Comparisons and Conclusions 12

3. Summary 13

  3.1 Summary of Conceptual Solution 13

  3.2 Technology Used 14

4. Software Requirements Specification 15

  4.1 Introduction 15

    4.1.1 Purpose 15

    4.1.2 Scope 15

    4.1.3 Glossary 15

    4.1.4 Overview of Document 16

  4.2 Overall Description 17

    4.2.1 Product Perspective 17

    4.2.2 User Characteristics 17

    4.2.3 Contraints 17

    4.2.4 Risks 17

    4.2.5 Assumptions and Dependencies 18

  4.3 Requirements 18

    4.3.1 Specific Requirements 18

    4.3.2 Functional Requirements 18

    4.3.3 Software System Attributes 19

4.4 UML ANALYSIS MODEL 20

    4.4.1 Use Case 20

    4.4.1.1 Brief Description Of Use Case Diagram 21

4.4.2 State Chart Diagram……………………………………………………………………22

5 Software Design Descripton……………………………………………………………23

5.1 Introduction 23

5.1.1 Purpose 23

5.1.2 Scope 24

5.1.3 Glossary.. 24

5.1.4 Overview of document 24

5.1.5 Motivation…………………………..…………………………………………..25

5.2 Design Overview 25

5.2.1 Description of Problem……………………………………………………………25

5.2.2 Technologies Used………………………………………………………………25

5.2.3 Architecture Design………………………………………………………………25

5.2.3.1 Sequence Diagram……………………………………………………………….25

5.2.3.2 Class Diagram…………………………………………………………………..27

5.2.3.3 Project Plan……………………………………………………………………..28

5.3 Data Design…………………………………………………………………29

5.3.1 Data Description………………………………………………………………………29

5.3.1.1 Entity Relationship (ER) Model……………………………………………..29

5.3.2 Data………………………………………………………………………………………30

5.4 Use Case Rezalition…………………………………………………………30

5.4.1 Project Components……………………………………………………………………30

5.4.1.1 Brief Description of Components…………………………………………….31

5.4.1.1.1 Ready Packages………………………………………………………………..31

5.4.1.1.2 Image Preprocessing…………………………………………………………..32

5.4.1.1.3 Image Segmentation…………………………………………………………32

5.4.1.1.4 Image Classification…………………………………………………………..33

6. Conclusions 34

7. References.. 34

Abstract

In today's world, after heart and lung diseases, cancer is the second biggest health issue that causes death. According to World Health Organization data; because of various cancer types particularly lung cancer; only in 2018, 9.6 million people lost their lives. In recent technological surveys that developed with latest technological tools, it is ascertained that detecting cancer and tumor on early stages has important value on extending patients lives. MR imaging is one of the technological tool used in medical that helps doctors to detect and see healthy and sick tissues. It has several advantages over other imaging techniques, providing high contrast between soft tissues. There has been various algorithms developed to process MR images. These algorithms gave different results depending on datasets and anatomic structures of images. Increasing computation of these algorithms and getting fattest, accurate results will make the patients start early treatment and will get ahead of possible bad outcomes. In this study, we aimed to improve developed algorithms to detect brain tumor area. This way we can help doctors to detect brain tumors in early ages. We did large reasearches on image processing algorithms that used on MRI ımages to detect brain tumor regions. After the research, it shows that Neural Network with back propogation or SVM with Hybrid method gives best results. Neural Networks need large datasets and more complicated. Since, we don't have any experiences on image processing, we decided to use and implement Hybrid Method with SVM. We will use TCIA's, kaggle's and BraTS's websites to find the datasets to use for this project. After doing a large research, we learned how most of the image processing algorithms that used for detecting brain tumor region and brain tumor type works. Such as K-means, Fuzzy, SVM, Neural Networks, K-NN, Region Growing.) We also learned that Python is best way to implement these methods and we gained a knowledge about how to implement these methods.

Key words:

Brain Tumor, MRI Images, Machine Learning, Image Processing, Artificial Intelligence, Deep Learning, Segmentation, Classification

1. Introduction

1.1 Problem Statement

Image processing is important subject used on medical innovations. Detecting brain tumor region is a critical and sensitive topic, because patients have more chance getting better or getting completely cured if they start early treatment. Therefore, early detection of tumors is very important. Mostly, image processing methods are used for detecting tumor region. There has been many experiments on detecting tumor region with image processing algorithms in past years. Even though, is has been found that some algorithms and methods works better than other. We did a broad research on this topic and we found out that, there hasn't been found one image processing algorithms that outperforms other algorithms. We are interested in image processing field and we wanted to improve our selves in that area. So we thought working on these experiments would be better for us beginners.

1.2 Background or Related Work

Mostly, other researches and students, professors who are interested in this field, made experiments on these subject and reliased articles in past years. The image processing methods can be used on and related to many other fields such as medicine, artificial intelligence, computer vision. Therefore, there is a lot of researches, articles and project related to this field. Our project is mostly related on image processing in medicine. Future experiments on this field could use our project and our articles as reference to our project and use our project to develop it. Also doctors who wants to get a second judgement on their diagnose can use our system.

1.3 Solution Statement

After doing research, we found out that Neural Networks or SVM with Hybrid Method gives the best result. We decided to use SVM with Hybrid method on this project. For Hybrid method, we decided to use K-means and Fuzzy algorithm. We decided to take some of the articles we found to consideration and implement them as reference. We will implement the methods used in these articles and try to improve them if possible. We will implement these methods on python programming language.

1.4 Contribution

We will combine some of the articles we found on our research. After combining the articles. The algorithms used on articles will work better and will give better results. Since, it is proven that Hybrid Method works better with past experiments. We also plan to develop these methods based on accuracy and computation time.

2. Literature Search

2.1 Examples of Segmentation Algorithms

Tresholding: Tresholding is one of the frequently used method for image segmentation. This method is effective for images with different intensities. Using this method, the image is partitioned directly into different regions based on the intensity values [11]. Edge based segmentation: Edge based segmentation methods partition an image based on rapid changes in intensity near edges [29]. The result is a binary image [11].

Region Growing Technique: Region growing method is one of the popular segmentation methods. This method starts with a seed pixel and grows the region by adding the neighboring pixels based on threshold value. When the growth of a region stops, another seed pixel which does not belong to any other region is chosen and the process is repeated [14]. The region growing is stopped when all pixels belongs to some region. Region growing segmentation is particularly used for delineation of small, simple structures such as tumors and lesions [15][11]. Region-growing approaches exploit the important fact that pixels which are close together have similar gray values.
Start with a single pixel (seed) and add new pixels slowly

1. Choose the seed pixel
2. Check the neighboring pixels and add them to the region if they are similar to the seed
3. Repeat step 2 for each of the newly added pixels; stop if no more pixels can be added [13].

The various limitations in using this method are

- Sometimes, manual interaction is required to select the seed point.
- Sensitive to noise so it produces holes or over segmentation in the extracted regions [16][11].
- The main assumption of this approach is that regions are nearly constant in image intensity.
- This approach will not work for non-smoothly varying regions (e.g., textured regions).
- More sophisticated approaches are required to allow for more realistic intensity variations due to shading [13].

Genetic Algorithm: (GA) is a meta-heuristic algorithm which is inspired by natural selection. Genetic algorithms provide solutions to search and optimization and problems. These algorithms depend upon bio-inspired operators. These operators are crossover, mutation, and selection etc apart from defining the optimization function [36].

Classifiers:

- **KNN K-Nearest Neighbor (k-NN):** K-NN classification technique is the simplest technique that provides good classification accuracy [17]. The k-NN algorithm is based on a distance function and a voting function in k-Nearest Neighbours, the metric employed is the Euclidean distance [18]. The k-NN has higher accuracy and stability **Step1:** Determine k value
- **Step2:** Calculate distance between the query instance and the training samples

- **Step3:** Sort the distance based on the kth minimum distance
- **Step4:** Assign the majority class
- **Step5:** Determine the class
- **Step6:** Segment the brain abnormalities [11]

- **SVM (Support Vector Machine):** Instead of minimizing an objective function based on the training samples (such as mean square error), the SVM attempts to minimize the bound on the generalization error (i.e., the error made by the learning machine on the test data not used during training). As a result, an SVM tends to perform well when applied to data outside the training set. SVM achieves this advantage by focusing on the training examples that are most difficult to classify. These "borderline" training examples are called support vectors. The SVM classifier is based on the hyperplane that maximizes the separating margin between the two classes. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. To maximize the margin between the classes its kernel is used to control the empirical risk and classification. There are many kernel functions such as linear, polynomial of degree and Radial basis function (RBF). Among these kernel functions, a radial basis function proves to be better for MRI brain images [20]. SVM has two stages; training and testing stage. SVM trains itself by features given as an input to its learning algorithm. During training SVM selects the suitable margins between two classes. Features are labeled according to class associative with particular class. Artificial neural network has a few issues having local minima and number of neurons selection for each problem. In order to resolve this problem SVM occupies no local minima and overhead to neurons selection by initiating the idea of hyper planes [11].

Clustering: A Clustering is one of the most useful techniques in MRI Segmentation, where it classifies pixels into classes, without knowing previous information or training. It classifies pixels with highest probability into the same class. Clustering technique training is done by using pixel features with properties of each class [21][11].

- **K-Means:** K-means clustering algorithm is the simplest unsupervised learning algorithm that can solve clustering problem. The procedure followed to classify a given set of data through a certain number of clusters is very simple. In K-means 'K' centres are defined, one for each cluster. These clusters must be placed far away from each other. The next step is to take a point belonging to a given data set and associate it to the nearest centre. When no point is pending, the first step is completed and early grouping is done. The second step is to recalculate 'k' new centroids as barycentre of the clusters resulting from the previous step. After having 'K' new centroids a new binding has to be done between the same data set points and the nearest new centre. A loop has been generated. As a result of this loop, the k centres change their location step by step until centres do not move any more.
  Algorithmic steps for K-means clustering:
  Let X = {x1, x2, x3, ...xn} be the set of data points and V = { v1, v2, v3, ...vc } be the set of centres.

  - **Step1:** Randomly select 'c' cluster centres

- **Step2:** Calculate the distance between each data point and cluster centres.
- **Step3:** Assign the data point to the cluster centre whose distance from the cluster centre is minimum of all the cluster centres.
- **Step4:** Recalculate the new cluster centre using

$$v_i = \left(\frac{1}{c_i}\right)\sum_{j=1}^{c_i} x_i$$

where 'Ci' represents the number of data points in ith cluster.
- **Step5:** Recalculate the distance between each data point and newly obtained cluster centres.
- **Step6:** If no data point was reassigned then stop, otherwise repeat from step 3. K-means algorithm is fast, robust and easier to understand. It also gives better result when data set are well separated from each other. But, if there are 2 highly overlapping data then k-means will not be able to resolve that there are 2 clusters [11].

- **Fuzzy C-Means:** FCM clustering is an unsupervised method for the data analysis. This algorithm assigns membership to each data point corresponding to each cluster centre on the basis of distance between the cluster centre and the data point. The data point near to the cluster centre has more membership towards the particular centre. Generally, the summation of membership of each data point should be equal to one [11].
Algorithmic steps for fuzzy C-means clustering:
Let X = {x1, x2, x3, ...xn}be the set of data points and
V = { v1, v2, v3, ...vc}be the set of cluster centres.

  - **Step1:** Randomly select 'c' cluster centres
  - **Step2:** Calculate the fuzzy membership 'µij' using the equation

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{ik}}\right)^{\left(\frac{2}{m-1}\right)}}$$

  - **Step3:** Compute the fuzzy centres 'vj' using

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{ik}}\right)^{\left(\frac{2}{m-1}\right)}}$$

- **Step4:** Repeat Step2 and Step3 until the minimum 'J' value is achieved or

$$\|U^{(k+1)} - U^{(k)}\| < \beta$$

where;

mentioning the vast amount of detail that each of those functions requires. Sometimes the function summary that is necessary for this part can be taken directly from the section of the higher-level specification (if one exists) that allocates particular functions to the software product.

'k' is the iteration step

○

'β' is the termination criterion between [0,1]

'U=(μ_{ij})n*c' is the fuzzy membership matrix

'J' is the objective function

The first loop of the algorithm calculates membership values for the data points in clusters and the second loop recalculates the cluster centres using these membership values. When the cluster centre stabilizes the algorithm ends. The FCM algorithm gives best result for overlapped data set and also gives better result than k-means algorithm. Here, the data point can belong to more than one cluster centre.

The main drawback of FCM is:

- The sum of membership value of a data point xi in all the clusters must be one but the outlier points has more membership value. So, the algorithm has difficulty in handling outlier points.

- Due to the influence of all the data members, the cluster centres tend to move towards the centre of all the data points [22]. It only considers image intensity thereby producing unsatisfactory results in noisy images [Hall et al. (1992)]. A bunch of algorithms are proposed to make FCM robust against noise and in homogeneity but it's still not perfect.

- **Adaptive Fuzzy Clustering:** The adaptive fuzzy clustering algorithm [22] is a modified version of standard FCM. The membership values in this method are calculated using equation:

$$\mu_{ij} = \frac{n * \left(\frac{1}{d_{ji}}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^{p}\sum_{z=1}^{n}\left(\frac{1}{d_{kz}}\right)^{\frac{1}{m-1}}}$$

This algorithm is efficient in handling data with outlier points. In comparison with FCM algorithm it gives very low membership for outlier points [22].

Artificial Neural Network: Neural Network based segmentation is totally different from conventional segmentation algorithms. In this, an image is firstly mapped into a Neural Network. Where every Neuron stands for a pixel [24], thus image segmentation problem is converted into energy minimization problem. The neural network was trained with training sample set in order to determine the connection and weights between nodes. Neural network segmentation includes two important steps feature extraction and image segmentation based on neural network. Feature extraction is very crucial as it determines input data of neural network [25], firstly some features are extracted from the images, such that they become suitable for segmentation and then they were the input of the neural network. All of the selected features compose of highly non-linear feature space of cluster boundary. Neural network based segmentation have three basic characteristics:

1. High parallel ability and fast computing [24].
2. Improve the segmentation results when the data deviates from the normal situation [24].
3. Reduced requirement of expert intervention during the image segmentation process. However there are some drawbacks of neural networks based segmentation either, such as:
    - Some kind of segmentation information should be known beforehand.
    - Neural network should be trained using learning process beforehand [24].
    - Period of training may be very long, and overtraining should be avoided at the same time.

Based on the architecture, Artificial Neural Network can be grouped into 2 categories:

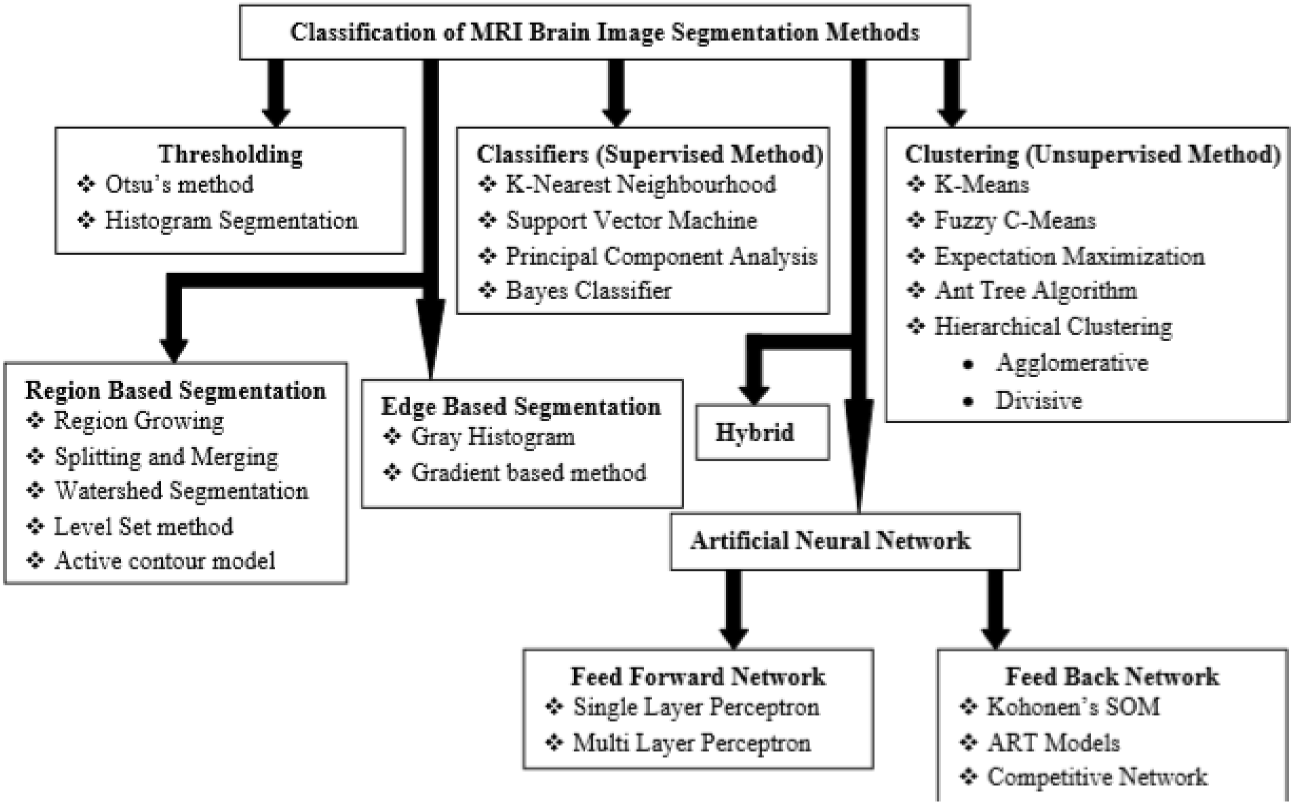
1. Feed-Forward Network
2. Feed-Backward Network or Recurrent Network

In feed-forward network, the neurons are arranged in layers that have unidirectional connections between them. Feed forward networks produce only one set of output values. It is called as static network because, the output values does not depend on previous output values. The output values are produced only based on current input. Feed forward network is also called as memoryless network. In feedback network, the neurons are arranged in layers that have bidirectional connections between them. Feedback networks produce a set of values which is updated based on the output values fed. Feedback network is also known as dynamic network because the output values depend up on the previous state values [11].

- **Back Propagation Algorithm:** Back Propagation algorithm [26] is used in layered feedforward ANN. This algorithm works for feed-forward networks with continuous output. In this network, the neurons are organized in layers and send their signals in the forward direction. The errors generated are propagated in the backward direction. The network receives the input by neurons in the input layer and the output of the network is given by the neurons on an output layer. The network consists of one or more intermediate hidden layers. Supervised learning is used in Back Propagation algorithm i.e., Examples of both the input and the output to be computed is given to the algorithm. The error between the input and the computed output is calculated. The network is trained with random weights and later the weights are adjusted to get the minimal error. The network will be perfect if the error is minimal. In back propagation, the weights and thresholds are changed each time an example is presented, such that the error gradually reduces. This is repeated until there is no change in the error [11]. MR image segmentation based on feed forward neural networks relies heavily on the training set used for their supervised training. The training set is constructed by selecting feature vectors from a single MR image or an ensemble of MR images and reflects the judgment of the human experts who assign labels to the feature vectors according to the tissues they represent [11][27].

Hybrid Techniques: Methods for performing segmentations vary widely depending on the specific application, imaging modality, body region and other factors. There is currently no single segmentation method that can yield acceptable results for every medical image. Combining several segmentation techniques together to form a hybrid framework can sometimes significantly improve the segmentation performance and robustness comparing to

each individual component. The hybrid segmentation approach integrates boundary-based and region-based segmentation methods that amplify the strength but reduce the weakness of both approaches
[28].



[11]

2.2 Related Literature

S. R. Telrandhe et al. [36] proposed brain tumor detection based on K-Means clustering algorithm that is unsupervised method based on neural network classifier. They used median filtering and morphological operation as preprocessing and it includes image filtering and skull masking. Preprocessed data was segmented with K-Means clustering and Object Labeling with Histogram of Oriented Gradients. In order to detect and classify brain tumor, Support Vector Machine (SVM) was used. Although K-Means clustering algorithm is easy to implement, accuracy is lower than expected. However, with the help of Object Labeling and SVM, detection of tumor can be easier and accuracy of classification can be raised.

L. Zhao & K. Jia [37] achieved considerable brain tumor segmentation accuracy by designing a three-stream framework named as multiscale CNNs that combines multimodal features from T1, T1-enhanced, T2, and FLAIR MRI images. Their proposed CNNs architecture was built with three-pathway streams: one way with small input image size 12 * 12, one with middle input image size 28 * 28, and last one with large size 48 *48. They used 2D CNNs to reduce computation time. Their model took the class of the central input patch as prediction result and input patch of the network is 2D patches from four modalities that are T1, T1c, T2 and FLAIR. They obtained output accuracy of about 90% for some dataset. CNN gives better results compare to K-Means, but its computation time and complexity may occur some problems. In order to achieve high accuracy results with CNN, our dataset should be as large and homogeneous as possible.

Random Forest approach, a solution for brain tumor detection, was discussed in [38]. These authors built forests of binary decision trees to avoid overfitting, using randomly chosen subsets of the training data, and randomly chosen subset of features for each tree. Histogram normalization, computed features and discarding missing data were used for preprocessing. They considered Dice score to compute accuracy. Initially the Dice score was 0.5339, however, they performed post-processing and it rose the Dice score to 0.8036. Their proposed methodology detects tumors of 2cm diameter. Creating binary decision trees and random forest is easy from the point of complexity. But, in order to reduce overfitting problem and increase accuracy, there should be lots of binary trees. In addition to that, with post-processing operation, computational time may be a problem.

S. Huda et al. [39] proposed a globally optimized hybrid significant tumor feature selection algorithm that is combined with bagging and decision tree to identify oligodendroglia’s with the chemo sensitive attribute and to minimize the effect of non-homogeneous or imbalanced data set. Their algorithm as called GANNIGMA (Globally Optimized Artificial Neural Network Input Gain Measurement Approximation) hybrid feature selection designed to find out significant features that help to create a simplified rule. Proposed hybrid approach integrated the knowledge about relationship between filter and wrapper approaches. This approach helped to provide a simplified decision rule and to obtain higher accuracy. Hybrid feature selection gives the highest accuracy results. However, it has the disadvantage of complexity, thus this may lead to difficulties at implementation steps. If there may occur a problem at some step of processing, stepping backward also causes wasting of time.

Proposed work [33] is aimed to develop an automatic method for brain tumor segmentation based on glowworm swarm optimization based fuzzy c-means clustering (GSOFCM) and region growing technique. The proposed method consists of three stages:

- Stage-1 is accelerating the FCM clustering for tissue segmentation process based on GSO.
- In Stage-2, is an abnormal detection process that helps to check the results of GSOFCM method by fuzzy symmetric measure (FSM).
- In Stage-3 is segment the tumor region from abnormal slices by region growing technique. GSO for selecting the initial centroids for the FCM method and segments the brain tissues such as background, gray matter (GM), white matter (WM), cerebrospinal fluid (CSF). Then abnormal detection process is done by FSM.
- Finally, the region growing method is applied in abnormal slices to segment the tumor region.
  The experimental results show that the proposed method gives satisfied results for brain tumor segmentation. The computation time for tumor segmentation ranged from 1 to 2 minutes depending on the size of the dataset. The study results compared to other methods is in the table below: [33]

S.No.	Methods	DC (%) (Complete Tumor)	Processing Time (min)
1	Geremia et al., 2012	62	10
2	Shin, 2012	30	8
3	Riklin Raviv et al., 2012	74*	8
4	Baurer et al., 2012	68	4-12
5	Buendia et al., 2013	57	21** and 20sec
6	Cordier et al., 2013	68	20
7	Festa et al., 2013	62	20-25
8	Taylor et al., 2013	44	1
9	Proposed Method	**75**	**1-2**

The proposed method resulted better at accuracy compared to other methods except [34] method and has faster computation time.

Proposed method [35] can not only localize the entire tumor region but can also accurately segment the intratumor structure. The proposed work was based on a cascaded deep learning convolutional neural network consisting of two subnetworks: (1) a tumor localization network (TLN) and (2) an intratumor classification network ITCN(IntraTumor Convolutional Network).The goal of the first subnetwork was to define the tumor region from an MRI slice. FCN-8s(Fully-Convolutional Network) is used for TLN.Then, the ITCN was used to label the defined tumor region into multiple subregions. Particularly, ITCN exploited a convolutional neural network (CNN) with deeper architecture and smaller kernel. The proposed approach was validated on multimodal brain tumor segmentation (BRATS 2015) datasets, which contain 220 high-grade glioma (HGG) and 54 low-grade glioma (LGG) cases. Dice similarity coefficient (DSC), positive predictive value (PPV), and sensitivity were used as evaluation metrics. Our experimental results indicated that our method could obtain the promising segmentation results and had a faster segmentation speed. More specifically, the proposed method obtained comparable and overall better DSC values (0.89, 0.77, and 0.80) on the combined (HGG + LGG) testing set, as compared to other methods [35].

All Magnetic Resonance Images (MRI) were normalized to 64x64 matrix of 10 mm slice thickness yielding a homogeneous element of 16 using Discrete Wavelet Transform (DWT) [40]. If demonstrates tumor, the neighboring 3 slices are analyzed to locate tumor margins and hence verify the prediction. If slices 6-7 and 11-12 were qualify is above , then all slices from 6-12 are considered segmentation. The left and the right sides of the brain are examined to determine if there is a tumor. They tested the algorithm with Principal Component Analysis (PCA) and without Principal Component Analysis (PCA), and found that there was no significant change in accuracy whereas the computation time increased increased significantly with its use.

Magnetic Resonance Imaging downloaded from medical image repository are present .mha format files [41]. MATLAB is not deal with .mha format files. Then it is converted to .jpeg format. Applied MR image Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA) technique for the extraction of features. After kernel Support Vector Machine classifier is built which is required for classifying images as normal or abnormal. This approach is a combination of DWT (Discrete Wavelet Transform) used for feature extraction, then the principal component analysis (PCA) for diminish the features and for the classification of MR images the Support Vector Machine has been used.

They remove the background of FLAIR image using the T1-weighted map [42]. For initial segmentation used Fuzzy C-Means (FCM) algorithm on intensity features. Fuzzy C-Means (FCM) algorithm segments into pre-specified number of clusters. Fuzzy C-Means (FCM) described the degree of similarity of one pixel each cluster. Compared the tumor region with T1-weighted and removed the mismatch regions. If slice does not contained any infarct region the they stopped the algorithm and declare slice is normal. This algorithm not take any prior assumption of shape and size of tumor regions. This frame work incorporate local as well as global information. This method incorporates additional information present in T1- weighted, which makes performance better in presence of artifacts and help to improves the boundaries.

In this study [43], they using MR images of the brain, segmented brain tissues into normal tissues such as white matter, gray matter, cerebrospinal fluid , and tumor-infected tissues. They used preprocessing to improve the signal-to-noise ratio and to eliminate the effect of unwanted noise. Furthermore, they used Berkeley wavelet transform. They compared 4 different methods. These methods K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), ANFIS, Back Propagation. They found Support Vector Machine (SVM) algorithms accuracy highest than K Nearest Neighbors (K-NN), ANFIS and Back Propagation. In this paper they used 4 different methods. In the result of experiments, SVM has the highest rate of specify, sensitivity and accuracy rate. K-NN and Back Propagation has the lowest specificity rate.

This work [44] aimed to prove that the consideration of the anatomical structure for the normal-abnormal classification will improve the result of the classification. The existing work shows that the feature vector, gray level co-occurrence matrix (GLCM) statistical features alongside Support Vector Machine (SVM) and Back Propagating Neural Network (BPNN) produce better results than other methods. Segmentation is done by using Multi Fractal. Then classification is done using two techniques Adaboost Classifier and machine learning based Back Propagating Neural Network. In Adaboost classifier the tumor is classified using Intensity Features where as in BPNN, classification is done based on the tumor area. Comparing to Adaboost, MLBPNN has an additional advantage of detecting whether the tumor is in early stage or in advanced stage. During this analyse process, the machine learning based Back Propagation Neural Network utilizes 30 sample images for making the tumour classification process.

- Accuracy: Adaboost: 67.15 / MLBPNN: 93.3
- Sensitivity: Adaboost: 68.18 / MLBPNN: 71.42
- Specifity: Adaboost: 53.04 / MLBPNN: 88.88

Comparing to Adaboost, MLBPNN has an additional advantage of detecting whether the tumor is in early stage or in advanced stage.

In this work [45], they proposed a model that includes the template-based K means and improved fuzzy C means (TKFCM) algorithm for detecting human brain tumors. In this proposed algorithm, firstly, the template-based K-means algorithm is used to initialize segmentation significantly through the perfect selection of a template, based on gray-level intensity of image; secondly, the updated membership is determined by the distances from cluster centroid to cluster data points using the fuzzy C-means(FCM) algorithm while it contacts its best result, the improved FCM clustering algorithm is used for detecting tumor position by updating membership function that is obtained based on the different features of tumor image including Contrast, Energy, Dissimilarity, Homogeneity, Entropy, and Correlation. Simulation results show that the proposed algorithm achieves better detection of abnormal and normal tissues in the human brain under small detachment of gray-level intensity. When they compare TKFCM with other algorithms,

- Sensitivity: TKFCM 97.43% | Thresholding 76.9% | ANN9 5.5%
- Computation time: Tresholding 3 min | TKFCM 40-50 s
- TKFCM know 38 abnormal image correctly in 40 data image.

Improved Fuzzy C-Means Algorithm gets better results with template based K-Means Algorithm when we compare them with other algorithms.

This work [46] describes the method of MRI brain image segmentation using Hierarchical self organizing map (Hsom). In this work they improve segmentation method consisting of two phases. In the first phase, the MRI brain image is acquired from patients' database, In that film, artifact and noise are removed after that HSom is applied for image segmentation. The HSom is the extension of the conventional self organizing map used to classify the image row by row. In this lowest level of weight vector, a higher value of tumor pixels, computation speed is achieved by the HSom with vector quantization. They calculated the number of tumor cells of different neighborhood pixel of 3 × 3, 5 × 5, 7 × 7, 9 × 9, 11 × 11 windows. Execution time of them is 13.76 in 795 number of segmented pixel when winning neuron 209, which makes this tool better than other segmentation tools. The target area is segmented and the evaluation of this tool from the doctor, whom the project is cooperated with, is positive and this tool helps the doctors in diagnosis, the treatment plan making and state of the tumor monitoring.

Twenty state-of-the tumor segmentation algorithms were applied to a set of 65 multi-contrast MR scans of low- and high-grade glioma patients and to 65 comparable scans generated using tumor image simulation software [47]. (BRATS) Different algorithms worked best for different sub-regions (reaching performance comparable to human inter-rater variability), but that no single algorithm ranked in the top for all sub-regions simultaneously. In whole tumor sensivity and specificity is better in Zahao method than other methods, but performance of Bauer is better than others. This benchmark cannot answer the question of what algorithm is overall "best" for glioma segmentation. They found that no single algorithm among the ones tested ranked in the top five for all three subtasks. Hamamci and Geremia performed comparably in the "off-site" and the "on-site" challenges, while the other algorithms performed significantly better in the "off-site" test than in the previous "on-site" evaluation. No single method performed best for all tumor regions considered. Their performance and sensitivity is changed from side to side.

2.3 Comparisons and Conclusions

- Edge detection methods require a balance between detecting accuracy and noise immunity. If the level of detecting accuracy is too high, noise may bring in fake edges making the outline of images unreasonable and if the degree of noise immunity is too excessive [24], some parts of the image outline may get undetected and the position of objects may be mistaken. Thus, edge detection algorithms are suitable for images that are simple and noise-free as well often produce missing edges or extra edges on complex and noisy images [30][11].

- Compared to edge detection method, segmentation algorithms based on region are relatively simple and more immune to noise [24]. Edge based methods partition an image based on rapid changes in intensity near edges whereas region based methods such as region growing, watershed and snakes, partition an image into regions that are similar according to a set of predefined criteria [11].
- In region growing methodologies are not standard methods for validate segmentation; the main problem is quality of segmentation in the border of tumor. This methods are good for homogeneous tumor but not for heterogeneous tumor [31].
- Classification based segmentation algorithms (K-NN, SVM, Bayers) segments tumor accurately and produce good results for large data set but undesirable behaviors can occur in case of where a class is under represented in training data [31].
- Clustering algorithms (K-means, Fuzzy, Hierarchical clustering) performs very simple, fast and produce good results for non-noise image but for noise images it leads to serious inaccuracy in the segmentation [31]. This could be solved by using accurate pre-processing algorithms (filtering, skull masking) before segmentation.
- Neural networks performs little better on noise field and no need of assumption of any fundamental data allocation but learning process is the biggest disadvantage of it [31].
- BPN (Back Propogation Neural Network) classifier with increase in nodes gives fast and accurate classification that can be effectively used for segmenting MRI brain images with high level of accuracy and the log sigmoid function is the best activation function for the medical image recognition application [32].
- SVM can classify the brain image as normal or abnormal more accurately than SOM (Self-organizing maps) neural network [32].
- It is concluded that Hybrid methods are better than using just one method. BPN Neural network, SVM, Fuzzy, K-means can be used for Hybrid Segmentation.

3. Summary

3.1 Summary of Conceptual Solution

In this project, we want to implement a system that detects the tumor region and tumor's type using image processing algorithms. There is a lot of experiments that does the same thing but we want to combine few of the past experiments and develop them considering their accuracy and computation time. We decided to use k-means and fuzzy algorithms because they are easy to implement. K-means and fuzzy algorithms are segmentation algorithms which means they only detect the area/region that wants to be detected but they don't necessarily classify the environment. Meaning segmentation algorithms only finds the different regions but does not define the region. We need to implement a classification algorithm that classy the area as tumor area and not tumor area. Mostly used and most successful classification methods are Neural Networks and SVMs. We decided to use SVM method to classify the region. From the research we did, we found out that SVM with Hybrid Method also gives one the best result among others. We choose few articles as reference for implementation. We will firstly use preprocessing algorithms to normalize and reduce noise on image. Then we will implement k-means and fuzzy algorithms to detect tumor region and lastly we will use classification methods to classify the region. This methods and experiments are planned to develop to help doctors with their diagnoses. We also want to help doctors with early detection of tumors so that patients can start early detection. These systems will be second hand judgement for their diagnoses.

3.2 Technology Used

We will use Python programming language and Opencv for the coding. Python works best with image processing since it also has special libraries we can use for image processing algorithms.

We will use Visual Studio Code and Jupyter editors.

We will use MySQL server for storing data because it works better with python programming language than other database servers.

There is no other technical tool needed for this project.

4. Software Requirements Specification

4.1 Introduction

4.1.1 Purpose

In past few decades, brain tumor segmentation in Magnetic Resonance Imaging (MRI) has become an emergent research area in the field of medical imaging systems. Accurate detection of brain tumor plays a vital role in the diagnosis of brain cancer. Our project aims to develop a program which analyses the Magnetic Resonance Images of the patient under consideration and recognizes the tumor by using image processing algorithms and detects the location of tumor. By using this method, patients can get information about their results . Purpose of this document is to describe and give details about the project that planned to develop a method that detects brain tumor from MR Images. Research students, professors , teachers and all kind of people that wants to do exploration about ımage processing segmentation,artificial intelligence neural networks can get benefit from this document. This document will also be helpfull, for researchers who will do further projects about this area in the future and wants to improve the methods done so far.

4.1.2 Scope

This system involves determining the location and type of tumor in MRI images. The algorithms used today are not sufficient enough. On latest experiments, it is shown that hybrid method is one of the most successfull and proper method for image segmentation. We intend to use Hybrid Method to detect whether or not brain tumor exists in MR Image and to detect brain tumor region. If the algorithm shows the tumor region properly, we plan to use another Hybrid method to detect tumor type. We want to show whether the tumor is cancerous or not. This way, the doctors will have second hand verification system for brain tumors. Algorithms will detect and highlight the tumor region area. If detection of tumor area is sucessfull as planned, the method will pass on to second part of the project and will detect tumor type as cancerous or not. So, the second part of the project is optional. This proposed method/system will help the doctors in diagnosing brain tumors, and help the patients to start early medication/treatment.

4.1.3 Glossary

Medical Imaging: Medical imaging is the technique and process of creating visual representations of the interior of a body for clinical analysis and medical intervention, as well as visual representation of the function of some organs or tissues

Image processing: Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image.

Image Segmentation: Image segmentation is the process of partitioning an image into multiple segments. Image segmentation is typically used to locate objects and boundaries in images.

Neural Networks: Neural networks are a set of algorithms, modeled loosely after the human brain, that are designed to recognize patterns.

Hybrid Method: Using more than one image processing algorithms on an image.

4.1.4 Overview

In the first part of this document we gave a brief description about the project and explained why this document is necessary. In the second part, we gave much more detail about the project, so it can be more understandable for readers and users of this method. Third part is much more detailed; this part gives software details of the project. Third part is mostly for users, researchers who wants to use this method on their own. Third part explains and gives helpful guide for them to use this method own their own. We gave detail about functional and non-functional features of this developed method.

4.2 Overall Description

4.2.1 Product Perspective:

The proposed method planned to work on both Linux and Windows operating systems. The method will work on brain MR Images only. The method will use the developed image processing algorithms already exists. We plan to use more than one image processing algorithms for detecting brain tumor area. We will use preprocessing for reducing noise on MR Images and use segmentation algorithms for detecting tumor area. The users of this method should have knowledge about medical structure of brain tumor or image segmentation or both. The code will be sharable and portable to the public after the method is published. The user can download the code on his/her computer and execute it for brain MR Images using any python editor.

4.2.2 User Characteristics:

The users of this method should have experience or knowledge about image segmentation and structure of brain and how it works. It is proposed that doctors in this area to use the method or research student and teachers who are interested in image processing and want to develop a project about this can do experiments with this method. It is recommended that doctors who use this method for their patients to not give the result of the method anyone other than their patients to prevent patient privacy. Users of this system should only use MR brain Images. User should know python programming language if they want to improve or test the proposed method.

4.2.3 Constraints:

The system only accepts Brain MR Images, if another image is given the system gives an error and does not work correctly. The system only works with specific operating systems. The system is coded in Python language. This code might not work with future version of python programming.

4.2.4 Risks

We did a research about medical image processing before starting this project. We plan to choose some of these articles we read and implement them, and combine the ones that gives the best result. The risk is that the implementation of these methods can not be fast enough so we may not start the second part of this project. Also, we will use different datasets from the experiments done in articles so the algorithms and efficiency might give different results depending on datasets. Finding enough amount of datasets and finding clean, accurate datasets are the biggest obstacles for this project

4.2.5 Assumptions and Dependencies

Future versions of python may not work with current version of python. The code of the method might have to change accordingly in the future.

4.3 Requirements

4.3.1 Specific Requirements

4.3.1.1. User Interfaces

There is no user interfaces for this Project since there is no website or application will be developed. The Project will be open source, and anyone who wants to access the code will be able to get phyton codes.

4.3.1.2 Hardware Interfaces

There is no hardware interface needed.

4.3.1.3 Software Interfaces

The codes will be available for both Linux and Windows. If the codes required for Linux and Windows Works differently,two different codes for both Linux and Windows will be shared.

4.3.1.4 Communications Interfaces

No communication interface needed.

4.3.2 Functional Requirements

User should be able to see the original MR Image,ımage after segmentation that shows the tumor region and image after detection tumor type.

The tumor region and type should be understandable for the user.

The code should be easily readable and understandable for the user.

The system should detect tumor region accurately and should detect less than a minute.

The system should implement the algorithm properly and accordingly as; first preprocessing algorithms then segmentation algorithms.

4.3.3 Software system attributes

I. Reliability

The method will be published after, we make sure the algorithms are developed properly and the method is detection the area %100 accurately and fast enough as planned.

II. Avaibility

The method will be available for Linux and Windows operating systems. The code will be published and accessable for everyone to download.

III. Performance

The system should detect tumor region accurately and should detect less than a minute.
The system should implement the algorithm properly and accordingly as; first preprocessing algorithms then segmentation algorithms.

IV. Portability

The system will be usefull in both Linux and Windows.

4.4. UML ANALYSIS MODEL

4.4.1 Use Case:

The following figure shows the use case diagram for the end users using the developed method. The most common users going to use the method will be doctors. This use case diagram shows the steps doctors can perform. Detecting Cancerous Tumor Region Use Case Diagram:

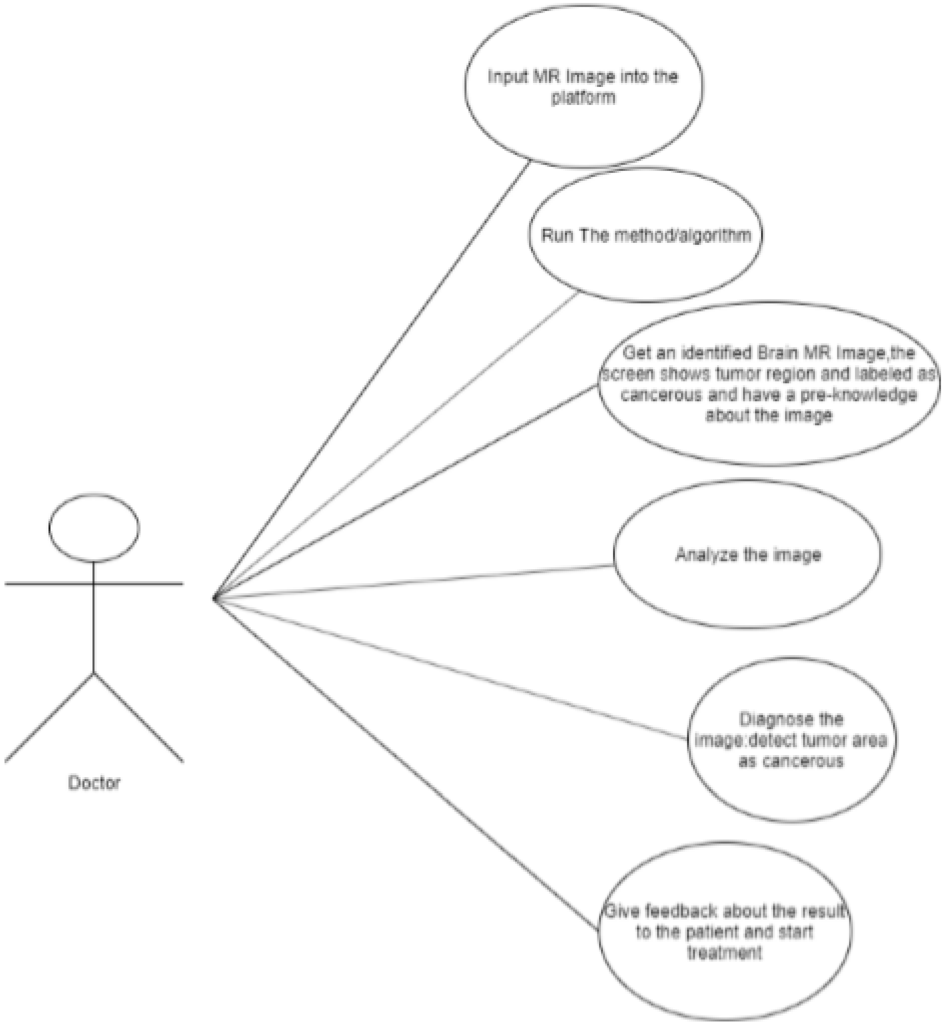


Figure 1: Use Case Diagram of proposed system

4.4.1.1 Brief Description Of Use Case Diagram:

If a patient is suspected with a brain tumor and a brain MR Image of the patient is taken. Doctors can input the achieved MR Image into the system and run the algorithm developed from this project. After some time, the result of the algorithms and segmentation will be showed in the screen, this way doctor can have a pre-knowledge before doing diagnoses to the image, the developed methods can help and reduce time for doctor's diagnoses. If there is no tumor found from the algorithms the screen won't show any tumor area. If there is a tumor found from the algorithm; the screen will show the region of the tumor and identify it as non-cancerous or cancerous. The doctor will give the patient his/her result of the MR Image.

Overview	
Title	Detecting Cancerous Brain Tumor Area in MR
Description	The doctor puts the MR Image of the patient to the system,systems shows that the MR Image has tumo area and has labeled it as cancerous.the doctor analyze the image to be sure and diagnose the area as cancerous.The doctors talks to the patient and starts treatment immediately.
Actors and Interfaces	Doctor,
Initial States and Preconditions	The inputted image should be only brain image and should be an MR Image.
Basic Flow	
.1.Doctor inputs image to the system 2.Run the Algorithm 3.After some little time,the screen shows an tumor area on the MR Image 4.The screen shows the tumor area as cancerous 5.Doctor analyzes the image 6.Doctor detects tumor area as cancerous 7.Doctor talks to the patient and starts the treatment	
Post Conditions	
None	
Alternative Flow(s)	
Detect Non-Cancerous Tumor Area: 1.Doctor inputs the MR Image into the system 2.Doctor runs the algorithm 3.The system detects a tumor area 4.The system labels cancer area as non-cancerous 5.Doctor analyzes the Image **Detect that there is no tumor:** 1.Doctors inputs the MR Image to the system 2.Doctors runs the algorithm 3.the system shows that there is no tumor 4.Doctor analyzes the Image	

Figure 2: Detect Brain Tumor Use Case Description

4.4.2 State Chart Diagram

State Chart of developed method:

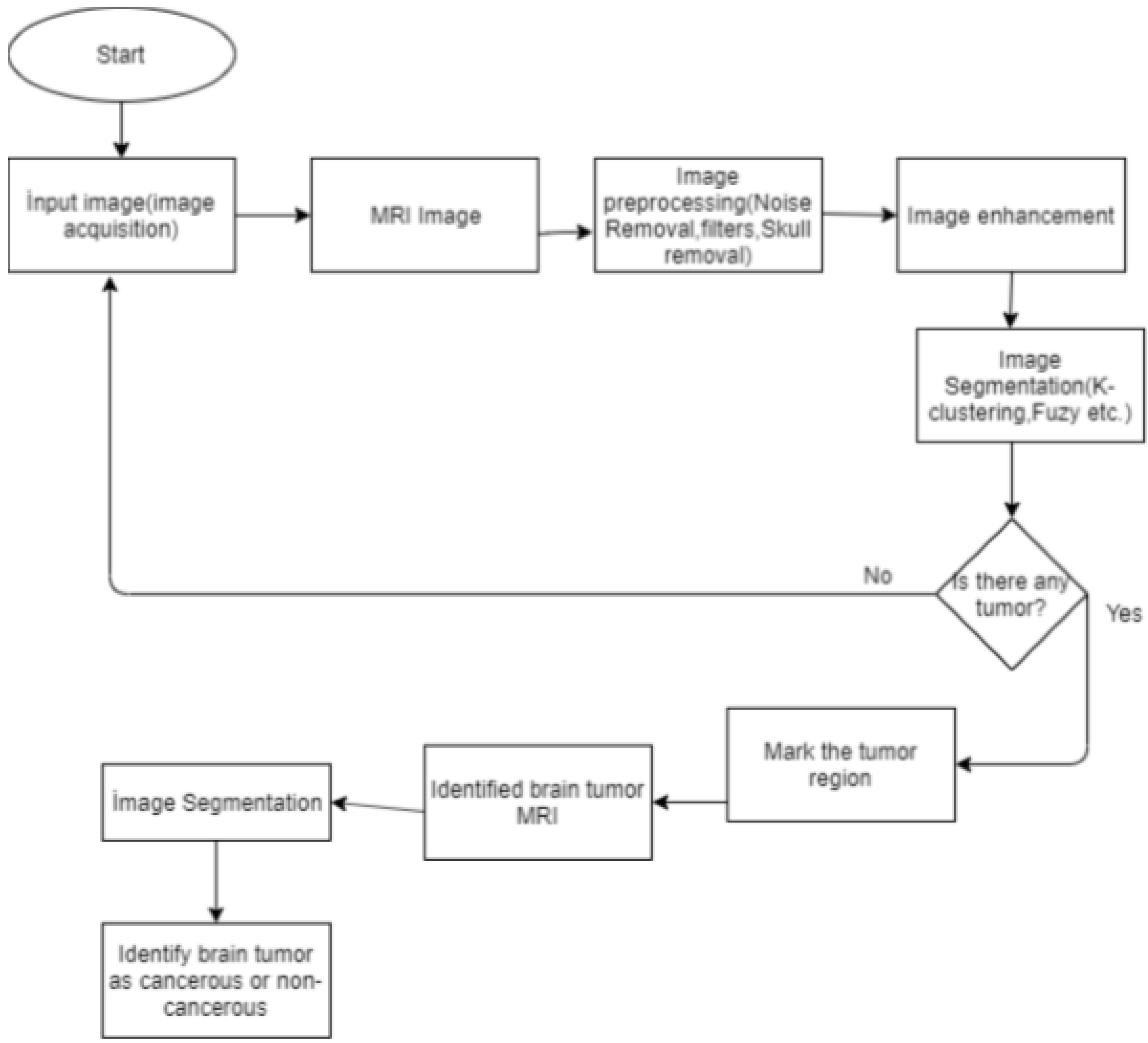


Figure 3: State Chart of proposed system

Flow chart of the system is planned to be closer to this example:

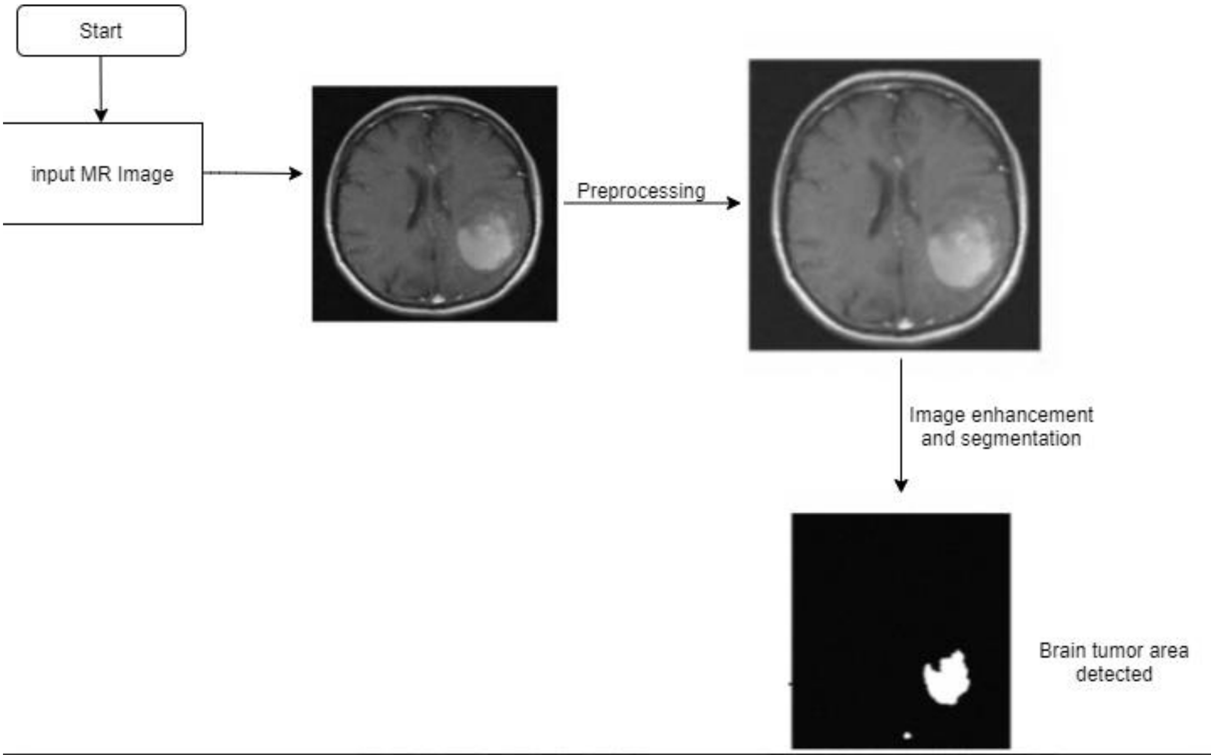


Figure 4: State Chart of reference system[36]

5. Software Design Description

5.1 Introduction

5.1.1 Purpose

This software design document describes the architecture and system design of a project that planned to develop a method that detects brain tumor location on MR Images. During the past few years, brain tumor segmentation in Magnetic Resonance Imaging (MRI) has become an emergent research area in the field of medical imaging systems. Accurate detection of brain tumor plays a vital role in the diagnosis of brain cancer. In past decades many experiments has been done in medical ımage processing and artificial intelligence areas for detecting brain tumor, but no one algorithm or method outperformed other methods. There is no algorithm found that works and detects the location of brain tumor %100 accurately and has a fast computation time. So, there is no one superior method. Accuracy and computation time are 2 important components for medical algorithms. Since, brain tumor is critical and sensitive topic, detection of tumor needs to be done properly and correctly. Over these experiments with algorithms, it is found that neural network methods and hybrid methods gives the best results considering computation and accuracy. Researchers , students ,professors and doctors who are interested in this field, wants to have more knowledge , wants to use or tests the method proposed can get benefits from this document.

5.1.2 Scope

For this project we want to work with medical images (MR Images) using image processing algorithms. We plan to use segmentation algorithms to detect whether brain tumor exists or not on brain MR Images. We plan to implement some of the methods used in past experiments in this field. We choose couple of experiments that gives the best results. We plan to implement these methods used on these experiments and compare their results depending on accuracy and computation time. We plan to combine these methods. From our research we chose few methods, we will combine these methods from the experiments. The combined methods are Fuzzy, K-means and SVM. Combining them will better results than using just one method. This is the first phase of the project. After first phase we will pass on to second phase and try to detect the type of the tumor as benign or malignant. We will also use preprocessing algorithms. MR Images carries a lot of noise which prevents the segmentation methods to work properly. These segmentation methods are used for detecting brain tumor region and noise decrease the efficiency of these methods working properly. So using preprocessing methods are necessary for this project. By using this method, doctors can get of their diagnosis verification and pre knowledge about the images. This project will also give contribution to medical image processing and artificial intelligence areas.

5.1.3 Glossary

Magnetic Resonance Imaging: Magnetic resonance imaging is a medical imaging technique used in radiology to form pictures of the anatomy and the physiological processes of the body.

Brain Tumor Segmentation: Brain tumor segmentation consists of separating the different tumor tissues (active tumor, edema and necrosis) from normal brain tissues.

Image processing: Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image.

Image Segmentation: Image segmentation is the process of partitioning an image into multiple segments. Image segmentation is typically used to locate objects and boundaries in images.

Neural Networks: Neural networks are a set of algorithms, modeled loosely after the human brain, that are designed to recognize patterns.

Hybrid Method: Using more than one image processing algorithms on an image.

Artificial Intelligence: In computer science, artificial intelligence, sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans..

5.1.4 Overview of document

This document gives technical details of the project planned to detect tumor region on MR Images. The first part of this document gives an overall description about the project. In the second part we gave detailed explanation of development of the project, design of the Project and planned Schedule for this project. In third part we gave technical details of the method we will develop.

5.1.5 Motivation

We are bunch of senior students in computer engineer department who are interested in image processing. Image processing has many fields. We want to improve ourselves on medical image processing. There is no, one algorithm that outperforms other algorithms on segmenting the medical images. Medical Image processing is a challenging and educational area, so we want to develop ourselves in this field. We also want to give good contribution to the world. Doctors can use this algorithms to be helpful for them on detecting brain tumors and start early treatment. This algorithms will be helpful to doctors. We did very large research in this field before starting the Project, we also studied python programming language. Python is the most suitable language for image processing since it has special libraries for image processing. Therefore, we will use python programming for this project.

5.2. DESIGN OVERVIEW

5.2.1 Description of Problem

There has been many image processing algorithms developed and many experiments done with these algorithms. Efficiency and accuracy of these algorithms are improved with these algorithms over the years. But, we still have long way to go to find the most successful algorithms. Detection of required area is sensitive and critical subject in segmenting medical images. Accuracy and fast computation time is two important scales for these segmentation algorithms. Hybrid methods and neural network methods gave the best results during these experiments but still there is no algorithm or method that leaves out other methods. We plan to implement some of these algorithms and combine them, if it’s possible we also want to improve them. These methods will also help doctors in diagnosing brain tumors. Doctors will use these methods before diagnosing brain MR Images and will have pre knowledge about the image and also they will get a second judgment about the image. This way patients diagnose will be verified by a second person and the patients will start early treatment.

5.2.2 Technologies Used

We will use python programming language, we will use Opencv libraries. We plan to use Visual Studio Code and Jupyter editors. We will use MySQL Server for storing database. There is no other technical tool needed for this project.

5.2.3 Architecture Design

5.2.3.1 Sequence Diagram

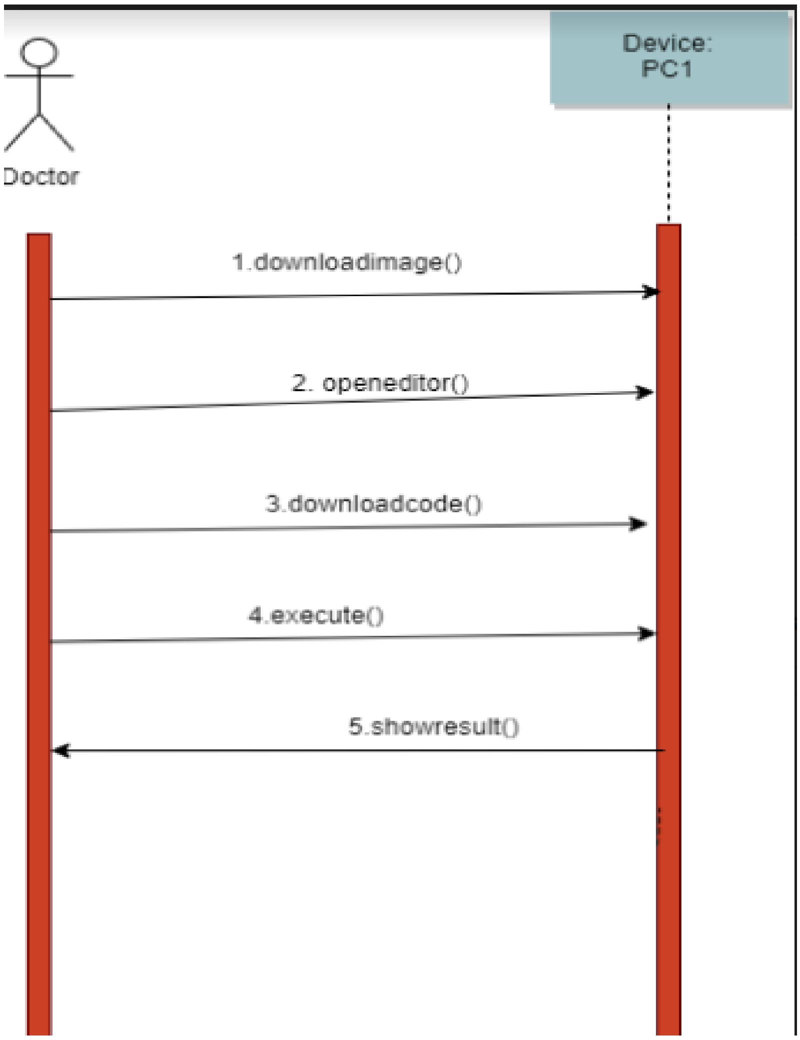


Figure 5: Sequence Diagram of proposed system

The Picture above shows the sequence diagram of proposed method. The user needs have an already existed brain MR image in their computer. The user then needs to download the code of our code and needs to open it in an editor that Works with python programming. The person than will execute the code and the system will show the results of the image. The system will Show the original image and image that is segmented (that shows the area of tumor).

5.2.3.2 Class Diagram

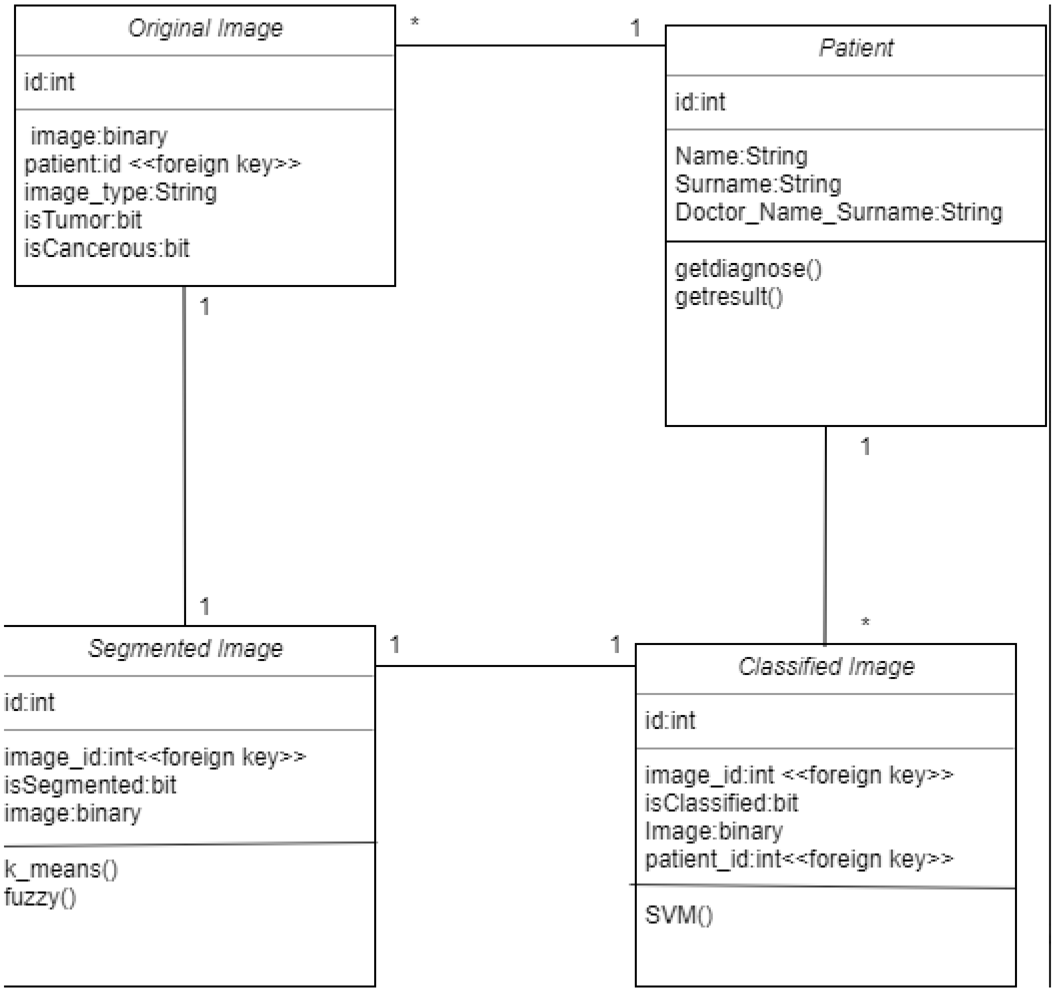


Figure 6: Class Diagram of proposed system

Class diagrams are used to show the classes, properties and relationships between each other in our project. There are 4 classes in our project. We decided to store original images, segmented images and classified images at different classes so that it can be easier to implement the algorithms and show them to the user. We decided to not store preprocessed images, since preprocessing is only used to reduce noise so that it can be easier to detect tumor region. Preprocessing is not directly related to detection of tumor. Therefore, we have 4 classes; Patient, Original Image, Segmentation Image, Classified Image. By Original Image, we mean original brain MR Images of the patients that has not been processed with any algorithms. The classes has id's as their priority keys, these priority keys defines different members of each class. The original image is directed to preprocessing than preprocessed image is directed to segmentation, then segmented image is directed to classification. This means that each original image has one segmented, classified and preprocessed image. Therefore, there is one-to-one relationship between the Original Images, Segmented Images and Classified Images. Classified ımages are results of the original ımages. Patient classes has one-to-many relationship with Original images and classified ımages since a patient can have more than one MR ımage.

5.2.3.3 Project Plan

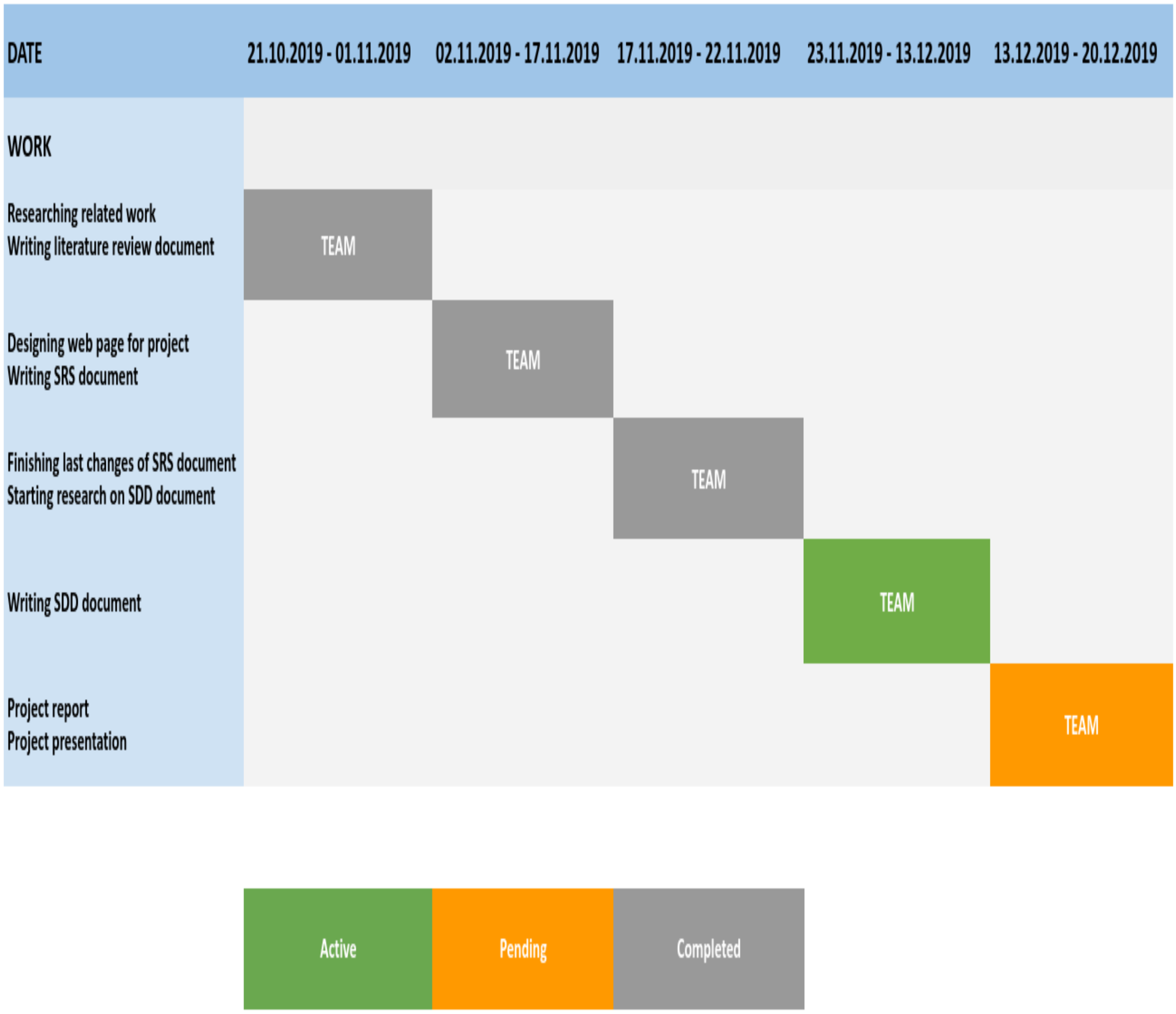


Figure 7:Table of Work Plan

Work	Description	Assignee(s)
Writing literature review document	Researching related literature, reading and documenting articles about the algorithms that will be used in the project	Team
Designing web page for project	Designing web page for publishing documents about project and team information	Team
Writing SRS document	Determining and documenting software requirements for project	Team

Work	Description	Assignee(s)
Writing SDD document	Identifying how data design and architecture design of software will be, documenting design description	Team
Project report	Collecting all documents about project in a single document	Team
Project presentation	Preparing presentation slide and making a demo presentation about what has been done so far	Team

5.3. DATA DESIGN

5.3.1 Data Description

5.3.1.1 Entity Relationship (ER) Model

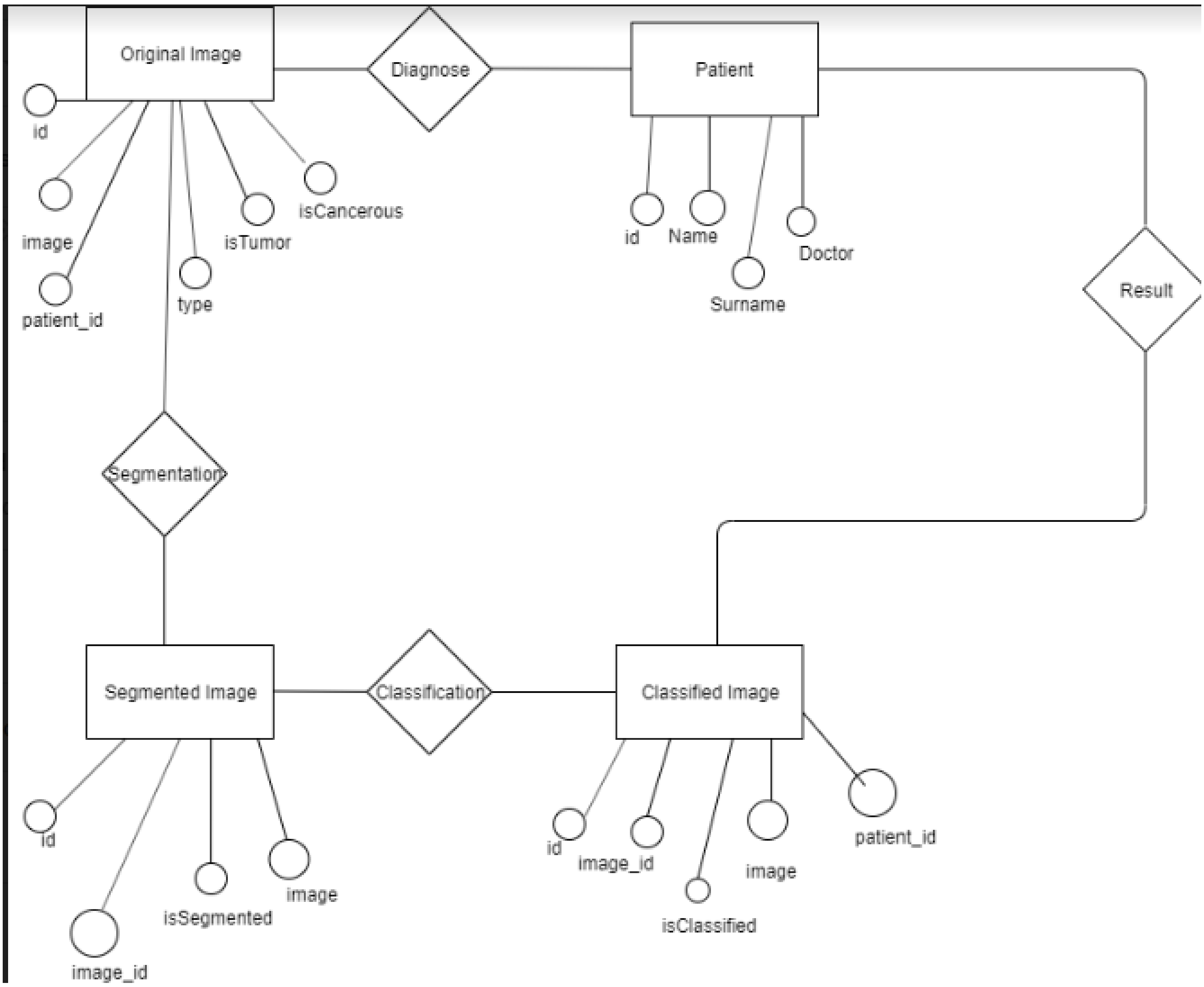


Figure 8: ER Model of proposed system

In the Figure, we showed the Image Analysis for the Classification of Brain Tumor Location on MR Images Entity Relationship Diagram. The ER Diagram shows Tables, Attributes, Relationships between tables will be used for database. The patients, original ımages, segmented images and classified images has id's as priority keys that defines them. Image_id attribute is a foreign key attribute for segmented and classified images that stores original image id. We store images as binary in these tables, so that it can be easier to show the images to users. We store type of the images on original image table. The brain MR Image's type is usually dicom but we will get the images from different sites so we want to also store them on type attribute on original image table. Original image has isTumor attribute as bit. It becomes 0 if there is no tumor detected on segmented image, it becomes 1 if there is tumor detected on segmented image. Original Image has, isCancerous attribute, it becomes 0 if there is no tumor detected or the tumor detected is found non-cancerous on classified image, it becomes 1 if detected tumor is found cancerous. We have isSegmented and isClassified attributes on Segmented ımage and classified image tables. isSegmented attribute becomes 1 after segmentation is complete, isClassified image becomes 1 after classification is complete. This way the system will be ensured to follow the steps accordingly and not to implement classification before segmentation.

We will use MySQL database server for storing database, since it is easy to use with python programming.

5.3.2 Data

The data to be used on the system should be brain MR images. MR image types are usually dicom and we expect to use dicom MR Images. Dıcom images usually stores the tissue of the image, type of the image and information about the patient in its format. DICOM (Digital Imaging and Communications in Medicine) is a standard for handling, storing, printing, and transmitting information in medical imaging. It includes a file format definition and a network communications protocol.[52]. We plan to get the MR images to use on the system and to implement algorithms from TCIA's[53], BRATS's[54] and kaggle's [55] websites.

5.4. USE CASE REALIZATION

5.4.1 Project Components

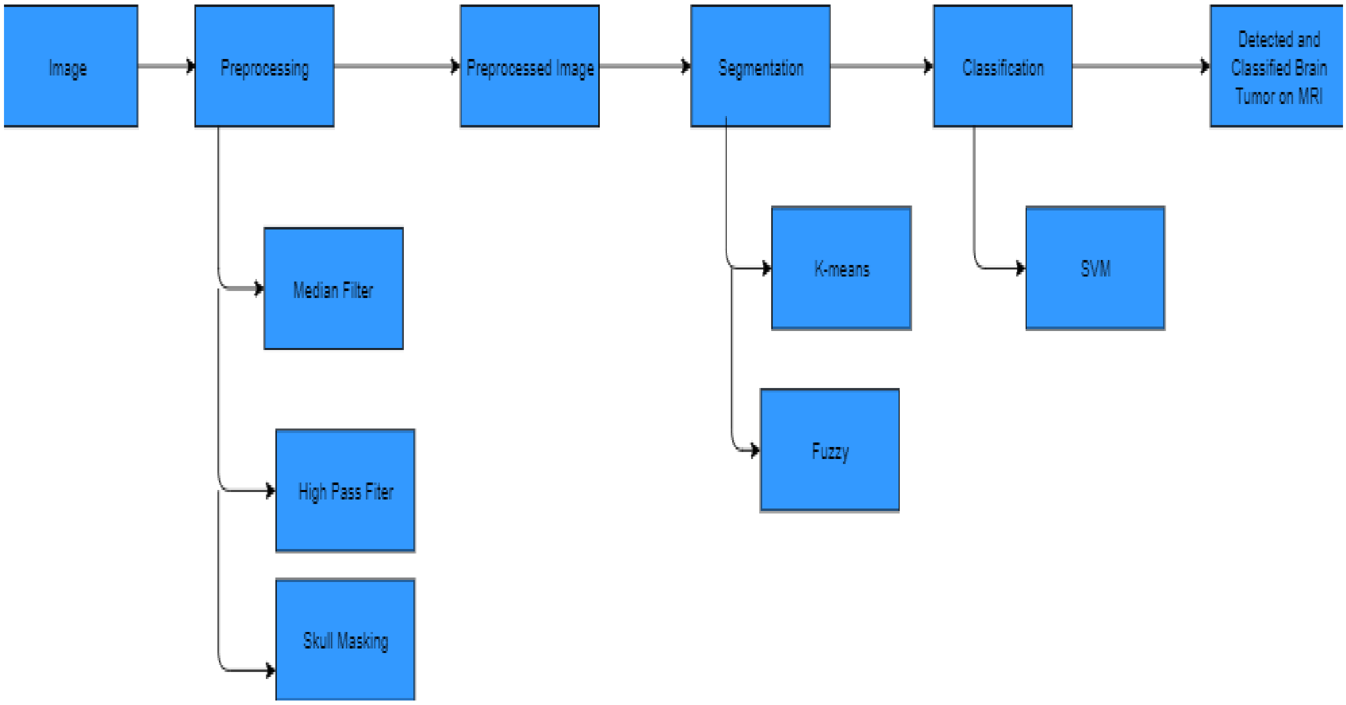


Figure 9: Block Diagram of Project Components

5.4.1.1 Brief Description of Project Components

From the research we did before starting this project, we picked 3 articles as references to implement for our project. These articles explain the details very well and very understandable and explains easy to implement algorithms. These articles mostly use k-means, fuzzy and svm. From the research we did, it shows that hybrid method with SVM shows almost as good result as Neural Networks. Therefore, we decided to use K-means and Fuzzy (Hybrid Method) and SVM. The articles we choose as references are: [36] [48] [47]

5.4.1.1.1 Ready Packages

Ready packages in Python which will be used for machine learning algorithms (clustering, classification) and for image processing are listed as below:

- Numpy: for multi-dimensional array and matrix processing (image).
- Scipy: for optimization, linear algebra, integration and statistics.
- Scikit-learn: for machine learning algorithms, supports most of the supervised and unsupervised learning algorithms (clustering and classification).
- Matplotlib: for data visualization.
- Pydicom: for dealing with MRI images.

5.4.1.1.2 Image Preprocessing

Pre-processing is a common name for operations with images at the lowest level of abstraction; both input and output are intensity images. The aim of pre-processing is an improvement of the image data that suppresses unwanted distortions or enhances some image features important for further processing (segmentation). [56] In this project, preprocessing algorithms are used to reduce noise and distortions so that, segmentation algorithms can detect tumor region more accurately. We plan to use 3 preprocessing algorithms for this project; Median Filter, High Pass Filter, Skull Masking.

Median Filter
The Median Filter is a non-linear digital filtering technique, often used to remove noise from an image or signal. Such noise reduction is a typical pre-processing step to improve the results of later processing. It is particularly effective at removing 'salt and pepper' type noise. The median filter works by moving through the image pixel by pixel, replacing each value with the median value of neighboring pixels. The pattern of neighbors is called the "window", which slides, pixel by pixel over the entire image. The median is calculated by first sorting all the pixel values from the window into numerical order, and then replacing the pixel being considered with the middle (median) pixel value. [57]

High Pass Filter
High pass filters are also called sharpening. A high-pass filter is a filter that passes high frequencies well, but attenuates frequencies lower than the cut-off frequency. Sharpening is fundamentally a high pass operation in the frequency domain. There are several standard forms of high pass filters such as Ideal, Butterworth and Gaussian high pass. Gaussian filter resulted better at sharpening the images filter in past experiments. A high-pass filter can be used to make an image appear sharper. [58]

Skull Masking
The quantitative morphometric studies of MR brain images often require a preliminary processing to isolate the brain from extra-cranial or non-brain tissues from MRI head scans, commonly referred to as skull stripping. Because the brain images that have preprocessed with automatic skull stripping eventually lead to get better segmentation of different brain regions which results for accurate diagnosis of various brain-related diseases. The brain regions must be skull-stripped prior to the application of other image processing algorithms. Moreover, skull stripping being a preliminary step, designed to eliminate non-brain tissues from MR brain images for many clinical applications and analyses, its accuracy and speed are considered as the key factors in the brain image segmentation and analysis. However, the accurate and automated skull stripping methods help to improve the speed and accuracy of prognostic and diagnostic procedures in medical applications. [59]

5.4.1.1.3 Image Segmentation

Image segmentation is regarded as an integral component in digital image processing which is used for dividing the image into different segments and discrete regions. The outcome of image segmentation is a group of segments that jointly enclose the whole image or a collection of contours taken out from the image.[60]

K-Means
K-means clustering algorithm is the simplest unsupervised learning algorithm that can solve clustering problem. The procedure followed to classify a given set of data through a certain number of clusters is very simple. In K-means 'K' centers are defined, one for each cluster.

These clusters must be placed far away from each other. The next step is to take a point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and early grouping is done. The second step is to recalculate ‘k’ new centroids as barycenter of the clusters resulting from the previous step. After having ‘K’ new centroids a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop, the k centers change their location step by step until centers do not move anymore. [36]

Fuzzy C-Means
FCM clustering is an unsupervised method for the data analysis. This algorithm assigns membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. The data point near to the cluster center has more membership towards the particular center. Generally, the summation of membership of each data point should be equal to one [36]

5.4.1.1.4 Image Classification

Image classification refers to a process in computer vision that can classify an image according to its visual content. For example, an image classification algorithm may be designed to tell if an image contains a human figure or not.[61]

SVM
Instead of minimizing an objective function based on the training samples (such as mean square error), the SVM attempts to minimize the bound on the generalization error (i.e., the error made by the learning machine on the test data not used during training). As a result, an SVM tends to perform well when applied to data outside the training set. SVM achieves this advantage by focusing on the training examples that are most difficult to classify. These “borderline” training examples are called support vectors. The SVM classifier is based on the hyperplane that maximizes the separating margin between the two classes. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. To maximize the margin between the classes its kernel is used to control the empirical risk and classification. There are many kernel functions such as linear, polynomial of degree and Radial basis function (RBF). Among these kernel functions, a radial basis function proves to be better for MRI brain images [36].
SVM has two stages; training and testing stage. SVM trains itself by features given as an input to its learning algorithm. During training SVM selects the suitable margins between two classes. Features are labeled according to class associative with particular class. Artificial neural network has a few issues having local minima and number of neurons selection for each problem. In order to resolve this problem SVM occupies no local minima and overhead to neurons selection by initiating the idea of hyper planes [36].

6. Conclusions

We learned details of image processing algorithms that used on medicine that we didn't know before. We learned mathematical details of these algorithms. We also had pre-knowledge of implementing these algorithms on python programming language. We gained more general knowledge about image processing field. We did a very broad research about the project. We read a lot of articles that detects brain tumor region on MRI Images using Image Processing methods. We planned the software requirements of the project and activity diagram of the project. We also planned the design of the project such as class diagram and database diagram. This way, we planned the project before starting the implementation to prevent any crisis and misleading. We plan to combine more than one experiment, this way the system will give better results than using just one method. We plan to develop the methods used in these experiments. The biggest risks of this project is finding accurate and enough dataset. Also we might not be able to develop and implement these methods properly. This is also one of the biggest risk of the project. This academic project will have contributions on image processing on medicine applications and future works. This project will also help doctors on detecting tumors on early stages.

7. References

[1] Kekre, H. B., Tanuja K. Sarode., Saylee M. Gharge (2009):"Detection and Demarcation of Tumor using Vector Quantization in MRI images", IJEST, Vol: 1, No: 2, pp: 59-66.

[2] MRI Basics- https://my-ms.org/mri_basics.htm

[3] T. Kalaiselvi, P. Nagaraja, "Brain Tumor Segmentation from MRI scans through Bit-plane Slicing and Clustering Techniques", In the Proceedings of the 2015 National Conference on Recent Advances in Computer Science and Applications (NCRACSA2015), Gandhigram, India, pp.89-94, 2015.

[4] B. Menze, A. Jakab , S. Bauer, J.K. Cramer, K. Farahani K, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)". IEEE Transactions on Medidal Imaging, Vol. 34, pp.1993-2024, 2015

[5] S.V. Aswathy, G. Glan Deva Dhas and S.S. Kumar, "A Survey on Detection of brain tumor from MRI brain images", IEEE International Conference on control, Instrumentation, Communication and computational technologies (ICCICCT), 2014.

[6] Luxit Kapoor and Sanjeev Thakur, "A survey on brain tumor detection using image processing techniques", IEEE international conference on cloud computing, data science, and engineering confluence, pp. 582-585, 2017.

[7] Rajesh C.Patel and A. S. Bhalchandra, "Brain tumor extraction from MR Images using Matlab", International Journal of electronic communication and soft computing science and engineering, vol. 2 no. 1, 2012.

[8] J. Sleva Kumar, A. Lekshmi and T. Arivoli, "Brain tumor segmentation and its area calculation in brain MR Images using K means clustering and fuzzy C means algorithm", IEEE international conference on advances in engineering science and management (ICAESM), 2012.

[9] M. N. Wu, C.C Lin and C.C. Chang, "Brain tumor detection using color-based K means clustering segmentation", Third International conference on intelligent information hiding and multimedia signal processing (IIH-MSP), 2007.
[10] N.N Gopal and M. Karnal, Diagnose brain tumor through MRI using image processing clustering algorithms such as fussy C means along with intelligent optimization techniques", IEEE International conference on computational intelligence and computing research, 2010
[11] MRI BRAIN IMAGES SEGMENTATION TECHNIQUES – A REVİEW, D.SELVARAJ Department of ECE, Sathyabama University, Chennai, Tamilnadu, India R.DHANASEKARAN, Syed Ammal Engineering College, Ramanathapuram, Tamilnadu, India, D.Selvaraj et.al / Indian Journal of Computer Science and Engineering (IJCSE), Vol. 4 No.5 Oct-Nov 2013
[12] TCIA(The Cancer Imaging Archieve) website - https://www.cancerimagingarchive.net/
[13] https://www.cse.unr.edu/~bebis/CS791E/Notes/RegionGrowing.pdf
[14] Mancas, M., Gosselin, B., Macq, B(2005): "Segmentation using a Region Growing Thresholding," Proceedings of the Electronic imaging Conference of the International society for optical imaging (SPIE/EI 2005), San Jose.
[15] Dzung L.Pharm, Chenyang Xu, Jerry L.Prince (1998): "A Survey of Current Methods in Medical Image Segmentation," Technical Report JHU / ECE 99-01, Department of Electrical and Computer Engineering.
[16] Brinkmann, B.H., Manduca, A., Robb, R.A(1998): "Optimized homomorphic unsharp masking for MR grayscale inhomogeneity correction," IEEE T. Med. Imag., 17:161–171.
[17] Warfield, S., Duda, (2001): "K-Nearest Neighbour Classification".
[18] El-Sayed, A., El-Dahshan, Abdel-Badeeh, M. S. Tamer, H. Y (2010): "A Hybrid Technique for Automatic MRI Brain Images Classification", digital Signal Processing, Vol. 20, Issue 2, March 2010, pp. 433-441.
[19] Clarke, L. P., Velthuizen, R. P., Phuphanich, S., Schellenberg, J. D., Arrington, J. A. Silbiger, M (1993): "MRI: Stability of Three Supervised Segmentation Techniques", Magnetic Resonance Imaging, 11: pp. 95-106, 1993.
[20] Zhang, K., CAO, H.X., Yan, H (2006): "Application of support vector machines on network abnormal intrusion detection", Application Research of Computers, Vol.5, pp.98-100.
[21] Pham, D., Xu, C. Prince, L. (2000):"Current methods in medical image segmentation", Annual review of Biomedical Engineering, PP. 315-337.
[22] Cox, E.,(2005): " Fuzzy Modeling and Genetic Algorithms for Data mining and Exploration," Elsevier, 2005
[23] Hall, L.O., Bensaid, A.M., Clarke, L.P., Velthuizen, R.P., Silbiger, M.S., Bezdek. J (1992): "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," IEEE Trans Neural Network, Vol.3, pp.672–682
[24] Kang, W.X., Yang, Q.Q., Liang, R.R (2009): "The Comparative Research on Image Segmentation Algorithms", IEEE Conference on ETCS, pp. 703-707.
[25] Zhu, C., Ni, C., Li, Y., Gu, G (2009): "General Tendencies in Segmentation of Medical Ultrasound Images", International Conference on ICICSE, pp. 113-117.
[26] Rumelhard, D., Mcclelland, J,(1986): "Parallel Distributed Processing", MIT Press, Cambridge, Mass.
[27] Jabbar, N., Mehrotra, M (2008): "Application of fuzzy neural network for image tumor description," Proc. of World Academy of Science, Engineering and Technology; 34:575-77.
[28] Chapter 1 Hybrid Segmentation Methods - https://academiccommons.columbia.edu/doi/10.7916/D8D226WB/download
[29] Gonzalez, Rafael C., Richard E.Woods (2007): "Digital Image Processing", Pearson Education.,

3rd Edition.
[30] Varshney, S. S., Rajpal, N., Purwar, R (2009): "Comparative Study of Image Segmentation Techniques and Object Matching using Segmentation", Proceeding of International Conference on Methods and Models in Computer Science, pp. 1-6.
[31] A Review on Automated Brain Tumor Detection and Segmentation from MRI of Brain, Sudipta Roy, Sanjay Nag, Indra Kanta Maitra, Samir Kumar Bandyopadhyay, 2013
[32] A REVIEW ON CURRENT MRI BRAIN TISSUE SEGMENTATION, FEATURE EXTRACTION AND CLASSIFICATION TECHNIQUES, D. SELVARAJ, R. DHANASEKARAN,2013
[33] Brain Tumor Segmentation from MRI Head Scans through GSO based FCM Clustering and Region Growing Technique, P.Nagaraja1, T.Kalaiselvi, P.Sriramakrishnan, JCSE,Vol-6, Special Issue-4, May 2018
[35] Automatic Semantic Segmentation of Brain Gliomas from MRI Images Using a Deep Cascaded Neural Network,Cui S,Mao L,Jiang JLiu C, Xiong S,J Healthc Eng, 2018 Mar
[36] S. R. Telrandhe, A. Pimpalkar, A. Kendhe (2016). Detection of Brain Tumor from MRI images by using Segmentation & SVM.
[37] L. Zhao, K. Jia (2016). Multiscale CNNs for Brain Tumor Segmentation and Diagnosis.
[38] Z. Kap´as, L. Lefkovits, L. Szil´agyi (2016). Automatic Detection and Segmentation of Brain Tumor Using Random Forest Approach.
[39] S. Huda, J. Yearwood, H. F. Jelinek, M. M. Hassan, G. Fortino (2016). A Hybrid Feature Selection With Ensemble Classification for Imbalanced Healthcare Data: A Case Study for Brain Tumor Diagnosis.
[40] T. Gupta, P. Manocha, T. Gandhi, R. Gupta, B. Panigrahi (2017). Tumor Classification and Segmentation of MR Brain Images
[41] S. Kumar, C. Dabas, S. Godara (2017). Classification of Brain MRI Tumor Images: A Hybrid Approach.
[42] C. Rao, P. Naganjaneyulu, K. Prasad (2017). Brain Tumor Detection and Segmentation Using Conditional Random Field
[43] N. Bahadure, A. Ray, H. Thethi (2017). Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM
[44] P.M. Shakeel, T.E. El. Tobely, G. Manogaran, H. Al-Feel (2019). Neural Network Based Brain Tumor Detection Using Wireless Infrared Imaging Sensor.
[45] M.S.Alam, Md Mahbubur Rahman, Mohammad Amazad Hossain, Md Khairul Islam, Kazi Mowdud Ahmed, Khandaker Takdir Ahmed, Bikash Chandra Singh, Md Sipon Miah (2019). Automatic Human Brain Tumor Detection in MRI Image Using Template-Based K Means and Improved Fuzzy C Means Clustering Algorithm .
[46] T. Logeswari, M. Karnan (2019). An Improved Implementation of Brain Tumor Detection Using Segmentation Based on Soft Computing.

[47] Mukambika P. S., Uma Rani K. IRJET. Segmentation and Classification of MRI Brain Tumor
[48] Mamidi Srujana , Mudadla Chandini , Ijjirythu Lavanya , R. Divyakanthi. Tumor Identification in Brain MRI Images Using Fuzzy-C-Means Algorithm

[49]
http://docwiki.embarcadero.com/RADStudio/Rio/en/Help_on_Help#Developer_Support_Services_and_Web_Site

[50] http://docwiki.embarcadero.com/RADStudio/Rio/en/Finding_information_(IDE_Tutorial)#Web-based_Product_Information

[51] https://web.cs.dal.ca/~hawkey/3130/srs_template-ieee.doc

[52] DICOM Library https://www.dicomlibrary.com/dicom/

[53] TCIA(The Cancer Imaging Archieve) https://www.cancerimagingarchive.net/

[54] BRATS(Brain Tumor Segmentation) http://braintumorsegmentation.org/

[55] kaggle https://www.kaggle.com/

[56] https://www.slideshare.net/ASHI14march/image-pre-processing

[57] https://www.cs.auckland.ac.nz/courses/compsci373s1c/PatricesLectures/Image%20Filtering_2up.pdf

[58] https://pdfs.semanticscholar.org/9b10/febd207e92f9d362ed06286d0ba31e290fec.pdf

[59] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4879034/

[60] https://www.sciencedirect.com/topics/computer-science/image-segmentation

[61]https://www.sciencedirect.com/topics/computer-science/image-classification

[62] IEEE Recommended Practice for Software Design Descriptions,2012, https://sovannarith.files.wordpress.com/2012/07/sdd_template.pdf

[63] IEEE Standard for Information Technology—Systems Design— Software Design Descriptions,2009 http://cengproject.cankaya.edu.tr/wp-content/uploads/sites/10/2017/12/SDD-ieee-1016-2009.pdf